This is an open access article published under an ACS AuthorChoice License, which permits copying and redistribution of the article or any adaptations for non-commercial purposes.

ACS central science

Research Article

Cite This: ACS Cent. Sci. 2019, 5, 1532−1540

http://pubs.acs.org/journal/acscii

Insights Into the Origin of Life: Did It Begin from HCN and H_2O?

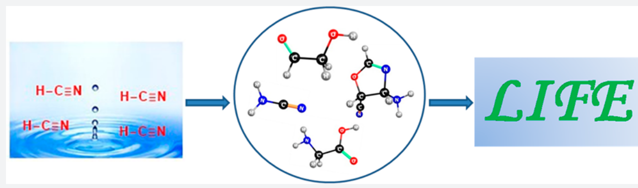
Tamal Das,[†,‡] Siddharth Ghule,[†,‡] and Kumar Vanka*,[†,‡]

[†]Physical and Materials Chemistry Division, CSIR-National Chemical Laboratory (CSIR-NCL), Dr. Homi Bhabha Road, Pashan, Pune 411008, India

[‡]Academy of Scientific and Innovative Research (AcSIR), Ghaziabad 201002, India

Supporting Information

ABSTRACT: The seminal Urey−Miller experiments showed that molecules crucial to life such as HCN could have formed in the reducing atmosphere of the Hadean Earth and then dissolved in the oceans. Subsequent proponents of the "RNA World" hypothesis have shown aqueous HCN to be the starting point for the formation of the precursors of RNA and proteins. However, the conditions of early Earth suggest that aqueous HCN would have had to react under a significant number of constraints. Therefore, given the limiting conditions, could RNA and protein precursors still have formed from aqueous HCN? If so, what mechanistic routes would have been followed? The current computational study, with the aid of the ab initio nanoreactor (AINR), a powerful new tool in computational chemistry, addresses these crucial questions. Gratifyingly, not only do the results from the AINR approach show that aqueous HCN could indeed have been the source of RNA and protein precursors, but they also indicate that just the interaction of HCN with water would have sufficed to begin a series of reactions leading to the precursors. The current work therefore provides important missing links in the story of prebiotic chemistry and charts the road from aqueous HCN to the precursors of RNA and proteins.



INTRODUCTION

How life originated[1−13] on Earth is one of the most fundamental questions of science, and has generated considerable interest. Research and discussion has resulted in two principal positions that are held today: the "RNA World" hypothesis[14−19] and the "metabolism-first" principle.[20−23] According to the RNA World hypothesis, life on Earth originated from the self-replicating molecules of ribonucleic acid (RNA),[24−26] which is the polymeric form of activated ribonucleotides.[27−29] The metabolism-first principle argues, on the other hand, that simple metal catalysts were present in the water in early Earth and aided in creating a soup of organic building blocks that subsequently formed the biomolecules necessary for life. The RNA World hypothesis has gained increased acceptance in recent times, with several experimental studies[26−28,30−35] indicating how hydrogen cyanide (HCN), known to exist on prebiotic Earth, could have been the starting point of many synthetic routes leading to the formation of RNA and protein precursors (see Figure 1A).

However, questions remain as to how HCN could have actually functioned in prebiotic conditions. As the famous Urey−Miller experiments have shown, HCN would have formed in the reducing atmosphere that existed during prebiotic times,[36] after which it would have condensed into the oceans.[13,37] HCN has a low boiling point, but at high pH (8−10), it is possible for it to exist in aqueous solution, even if the temperature of the water is 80.0−100.0 °C. However, since the hazy atmosphere[38] of the Hadean Earth would have made it difficult for high-energy photons to reach the Earth's surface (much like the red surface of Titan today, because of a similar haziness in the atmosphere), a lot of the reactions shown in Figure 1, which depend upon photochemistry or an electric spark, may not have been possible for aqueous HCN. Hence, the more plausible alternative would have been thermochemistry. It is possible that temperatures at the surface of the water bodies of early Earth (3.5−4.0 billion years ago) would have been about 80.0−100.0 °C,[39] which suggest favorable conditions for thermochemistry, but if thermochemistry predominated in the oceans of early Earth, it could be argued that hydrolysis would have taken precedence over the polymerization of HCN. This is because HCN polymerization would have had to begin with HCN dimerization and the subsequent reaction of the product with more HCN molecules. In other words, the polymerization of HCN would have required a series of second-order reactions in HCN, while the competing hydrolysis of HCN would have simply required the HCN collision with the surrounding solvent water molecules. Indeed, previous studies[34,40] have shown that in dilute aqueous concentrations of HCN, hydrolysis is favored over oligomerization.

Then, there is also the issue of too-high temperatures: experiments have shown[41,42] that at temperatures above 100.0 °C, decomposition of the formed RNA and protein precursors would occur. Therefore, the reactions would have had to happen around 100.0 °C,[43,44] which indicates that the barriers (ΔG values) of the reactions of monomeric HCN in water could have only been about 40.0 kcal/mol: previous computa-

Received: May 28, 2019
Published: August 7, 2019

 © 2019 American Chemical Society DOI: 10.1021/acscentsci.9b00520
ACS Cent. Sci. 2019, 5, 1532−1540

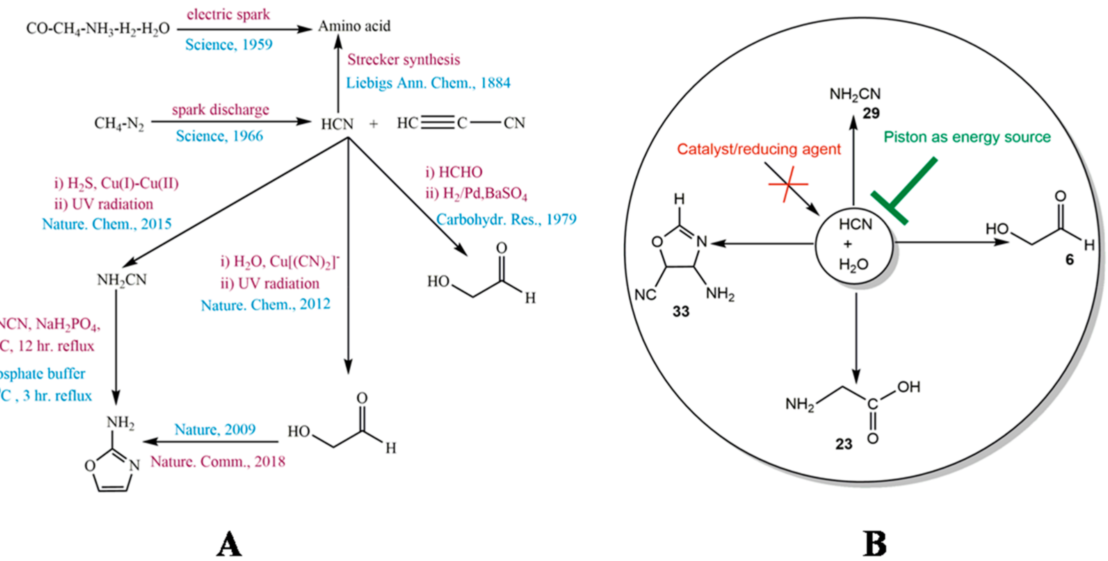


Figure 1. (A) Previously synthesized RNA and protein precursors (amino acids, cyanoacetylene, cyanamide, glycoaldehyde, and 2-amino-oxazole). (B) The ab initio nanoreactor (AINR) approach, yielding RNA and protein precursors, beginning from only two different reacting molecules, HCN and H_2O, obtained in "one-pot", under the same reaction conditions.

tional studies[45−49] have shown that chemical reactions occurring at temperatures of around 100.0 °C have barriers in the region of 40.0 kcal/mol. Barriers higher than 40.0 kcal/mol would have led to much slower reactions (or no reactions) at 100.0 °C. Slower reactions may be possible at higher barriers, but this would lead to the possibility of other side reactions also becoming competitive and causing significant reduction in the formation of desired products.

Hence, for the RNA World hypothesis to be true, there are several constraints that have to be kept in mind: (i) thermal, not photochemical conditions, (ii) reactions where monomeric and not polymeric HCN would predominate, (iii) without mediation from metal catalysts, (iv) at temperatures not exceeding 100.0 °C, and (v) having reactions with free energy barriers not exceeding 40.0 kcal/mol. To this list, one could add (vi) the need to avoid chemical processes involving the protonation of substrates, since HCN, with a pK_a of 9.31, is a weak acid and would have largely remained in undissociated form in solution. The protonation of water would also have been suppressed since the pH of water has been estimated to be between 8.0 and 9.0 in prebiotic times at the surface of the ocean.

This list of conditions appears formidable and leads to the important question: could life have begun under these circumstances? This current work attempts to answer this question, through the agency of the ab initio nanoreactor (AINR).

The AINR method, recently developed,[50] allows one to obtain reaction pathways and products without controlling the chemical system in any way.[51,52] This represents a major shift in what one can do with computational chemistry, because, using the AINR, one can now *discover* new reactions, completely independent of experimental input. This was demonstrated by Martinez and co-workers[50] when they found entirely plausible, new pathways for the formation of amino acids, from a computational re-enactment of the Urey−Miller experiment.[53] In the current work, we have conducted full quantum mechanical molecular dynamics (MD) simulations on systems employing the AINR approach on systems containing a mixture of molecules of HCN and H_2O. The goal has been to follow the chemical reactions that can occur through collisions between the molecules and observe what new species are formed as a result. In short, our objective has been to perform the equivalent of an experimental study while satisfying the conditions outlined in (i−vi) above. Remarkably, we have found, as will be shown in the Results and Discussion, that just the interaction of HCN and H_2O was sufficient to eventually lead to the formation of the experimentally reported precursor molecules to RNA and proteins:cyanamide,[27,30−32,54,55] glycolaldehyde,[27,30,31,56−61] an oxazole derivative,[27,62,63] and the amino acid glycine[36,53,64−68] (as shown in Figure 1B). Furthermore, analysis of the data allowed us to determine the mechanistic pathways by which HCN and H_2O reacted together to yield intermediates and, eventually, the RNA and protein precursors. We subsequently subjected these pathways to a full static quantum chemical study with density functional theory (DFT) and thus obtained all the barriers ($\Delta G^{\#}$) for the reactions involved in these processes, as well as the energies (ΔG) of the reactions. As will be discussed in the Results and Discussion, this has led to results that not only reveal interesting pathways for the formation of the precursor molecules beginning from aqueous HCN but also indicate that these mechanistic routes would have been thermodynamically and kinetically feasible.

■ RESULTS AND DISCUSSION

The AINR approach makes use of collisions between the molecules of HCN and H_2O, and this provides the energy required to cross the activation barriers for each of the elementary steps of the reactions. The simulations have been done on systems having nearly homogeneously mixed HCN and H_2O molecules as the starting reactants. Sixteen H_2O and 15 HCN molecules were taken together, and the system was allowed to evolve for 750 ps. For more details, please see the Computational Methods section. Collisions between the molecules gave rise to new species. It should be noted that homogeneous mixtures of HCN and H_2O do not represent the exact ratios of HCN and H_2O present in a localized region of

DOI: 10.1021/acscentsci.9b00520
ACS Cent. Sci. 2019, 5, 1532−1540

the early Earth oceans or hydrothermal vents. The reason such mixtures were employed was to maximize the possibility of interactions between HCN and H_2O in the AINR. This would increase the probability of obtaining different products during the simulations. The goal of the AINR simulations was to obtain mechanistic pathways for the formation of different RNA and protein precursors beginning from HCN and H_2O, pathways that could then be studied carefully with a static DFT and QM approach to ascertain their feasibility. Studying homogeneous mixtures of HCN and H_2O afforded the best possibility of realizing this goal.

Figure 2 below illustrates how a system starting with a mixture of HCN and H_2O molecules evolves with time. An mp4 file (Supporting Information, Movie S1) of a movie made of a part of an AINR simulation is included in the Supporting Information.

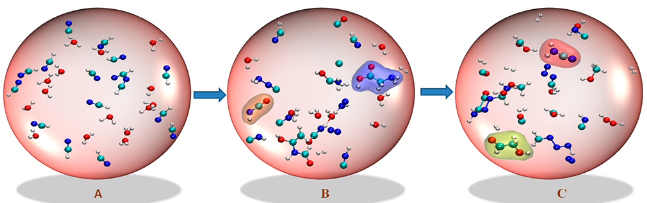


Figure 2. Snapshots of AINR simulations. (A) the beginning, 0.0 ps: only HCN and H_2O present. (B) after 100 ps, glycine (blue surface) has formed, along with molecules such as isocyanic acid (pale yellow surface). (C) after 250 ps: glycoaldehyde (green surface) and cyanamide (orange surface) have formed, along with other oligomeric species. Color scheme: oxygen: red, carbon: teal, hydrogen: gray and nitrogen: blue.

The AINR approach thus leads to the discovery of new species from the starting compounds, and the analysis of the data through the connectivity graph (shown in Figure S1, Supporting Information) allows the exploration of new mechanistic pathways. The next section discusses the results that have been obtained by this approach.

Analysis of the Reaction Pathways Leading to the Formation of Specific Compounds. *Formation of the Protein Precursor: Glycine.* The use of the nanoreactor produces as an output many different pathways to new species from the starting reactants. In most of the pathways, formaldehyde **3**, urea **26**, formaldimine **2**, and glycolonitrile **4** were seen to be formed as intermediates. This suggests that these species were the key intermediates en route to the formation of the target molecules, as has also been noted by experimentalists.[29,30,35,60,63,69,70] Apart from this, a lot of diverse organic species were also observed to have formed during the simulations (Figure S2). Moreover, small molecules such as CO **13**, CO_2 **11**, and H_2 were produced (Figure 3), and these were seen to take part in the synthesis of comparatively larger organic molecules. Figure 3 below describes how the relevant intermediates formic acid, CO, and CO_2 are formed from HCN and H_2O, leading from HCN, **1**, through the intermediate species **7**, formamide **8**, formic acid **9**, to carbon monoxide, CO, **13**. The complete free energy profile is shown in Figure S3, Supporting Information.

Figure 3 shows that species such as ammonia and dihydrogen were created in the AINR from the interaction between the HCN and water, and they turned out to be important substrates in subsequent reactions. This is interesting because it suggests that HCN and water would have created all the necessary reactants in subsequent steps. However, it could be argued that since the concentration of HCN would have been low such species would have been formed in low concentrations as well, which would have further reduced the yield of the subsequent products. The counter argument to this is that ammonia and other reactant species were also present separately in the oceans at that time because the protective haze of the Titan-like atmosphere would have prevented the photochemical degradation through UV of molecules such as ammonia in the atmosphere, and such molecules could have dissolved separately in the oceans as well and could be thus available for the reactions shown later in Figure 4.

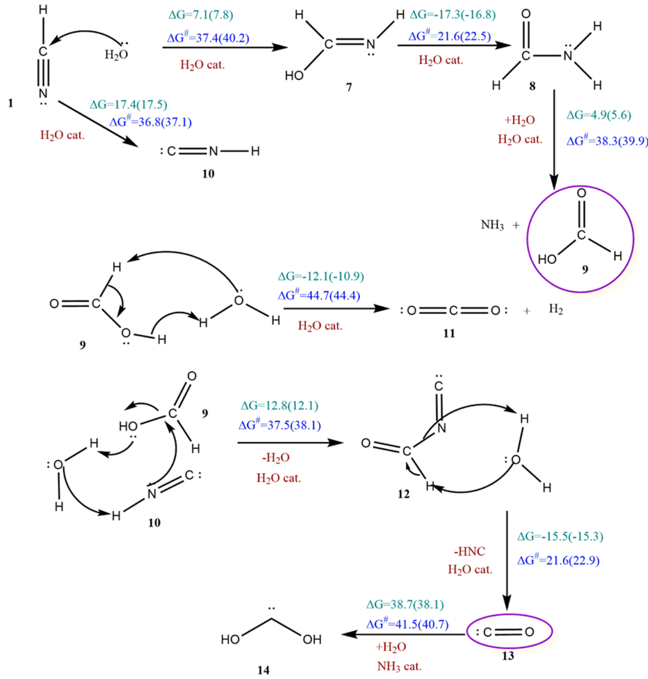


Figure 3. Sequence of elementary reaction steps derived from the AINR: the formation of HCOOH, CO_2, and CO starting from HCN and H_2O. Molecules labeled "cat.", shown in brown, participate catalytically as proton shuttles. Values have been calculated at the B3LYP-D3/TZVP+COSMO(ε = 80.0)//RI-CC2/TZVP+COSMO-(ε = 80.0) and the B3LYP-D3/TZVP+ COSMO(ε = 80.0)//RI-MP2/TZVP+COSMO(ε = 80.0) (values shown in parentheses) levels of theory in kcal/mol.

Now, as we continue along this path, we find that important intermediate species: formaldehyde **3** and formaldimine **2** are formed during glycine **23** synthesis via several elementary steps (see Figure 4A and energy profile in Figure S5). These intermediates lead to the formation of glycine **23**, the precursor to proteins. A perusal of the three most feasible pathways found for glycine formation (two are shown in Figure 4B and one in Figure S7, Supporting Information) shows that **2** is present as an intermediate in all the cases. All the pathways are seen to proceed via stepwise elementary reaction steps. In one of the pathways, **2** reacts with hydrogen cyanide **1** to produce **20**, which, through further stepwise hydrolysis, leads to **23**. This is the well-known Strecker synthesis pathway.[64] That the AINR finds the same is gratifying and can serve as a validation for the computational approach. What is also satisfying is that the pathway does not involve protonation, which therefore agrees with condition (vi), mentioned earlier. We do note, though, that the AINR has also found another pathway,

DOI: 10.1021/acscentsci.9b00520
ACS Cent. Sci. 2019, 5, 1532−1540

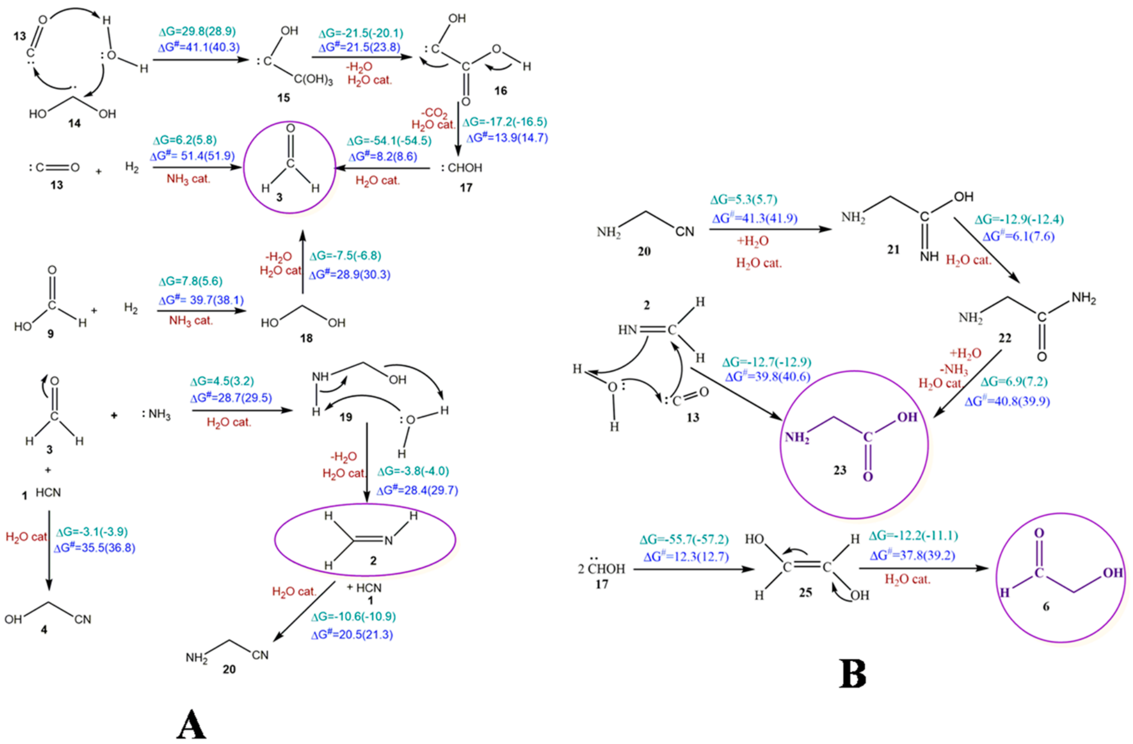


Figure 4. (A) The sequence of elementary reaction steps derived from the AINR: the formation of formaldehyde, formaldimine, glycolonitrile, and aminoacetonitrile. (B) The formation of the target species: glycine and sugar. The values are in kcal/mol.

involving a trimolecular reaction between **13**, **2**, and water, with a five-membered transition state, which also leads to **23** (encircled in violet in Figure 4B).

Pathways for RNA Building Units. The previous section discussed feasible pathways for the formation of the important protein precursor: glycine **23**. In this section, we discuss how the AINR also yields feasible pathways for the formation of important RNA precursors such as cyanamide **29**, glycoaldehyde (sugar) **6**, and the oxazole derivative **33**. The experimental synthesis of the sugar **6**, one of the most important RNA precursors, is challenging, but several reports have emerged recently where this has been achieved. Recently, Sutherland and co-workers[60] synthesized sugar from HCN and H_2O in the presence of a copper cyanide catalyst, through a photoredox cycle. They proposed a mechanistic pathway for the formation of the sugar **6** in the absence of the copper cyanide catalyst (Figure S8A, Supporting Information), involving two reduction steps, through which HCN **1** would be reduced to **2** and **4** to imine **5** in the presence of an H_2O catalyst molecule. We have calculated the barriers for these two processes (at the same level of theory as the mechanistic pathways investigated in the current work) and found them to be 83.2 and 85.8 kcal/mol respectively (see Figure S8B, Supporting Information). Interestingly, the results obtained from the current computational studies reveal a completely different pathway for sugar formation, avoiding the reduction steps. The feasible pathway derived from the nanoreactor for the formation of **6**, where two molecules of hydroxymethylene carbene **17**, the formation of which was discussed in the previous section (see Figure 4A), dimerize with a low barrier to form **24**, which further tautomerizes and leads to the formation of **6** (encircled in violet in Figure 4B). This pathway is found to be facile, with reduction not involved in any of the steps, with the slowest step seen to be ∼37.0 kcal/mol, which is considerably lower (by more than 40.0 kcal/mol) than the barriers for the pathways (83.2 and 85.8 kcal/mol) that have been proposed in the literature.[60] The implication of this is that the process of forming sugar **6** would not have needed the presence and intervention of metal catalysts but would have been possible under thermal conditions, at temperatures of 80.0−100.0 °C. For other pathways, please see Figure S9, Supporting Information.

It is also interesting to note that Schreiner's group has recently reported[70] that carbene **17** is the intermediate en route to the formation of **6**. The fact that the AINR discovers the same intermediate, carbene **17**, to be important for the formation of **6** as this recent, independent, experimental study by Schreiner and co-workers is quite remarkable. We note, however, that the pathway that has been proposed for the transformation of **17** to **6** in the absence of solvent and base by Schreiner and co-workers (shown in Figure S10, Supporting Information) is different from the one that we have discussed here. Since both the solvent and the base are relevant in our simulations (and in early Earth), it is quite likely that the facile pathways reported here transforming **17** to **6** represent avenues by which this important RNA precursor was formed in prebiotic Earth.

Cyanamide **29**, another precursor of RNA, is formed via two important intermediate species isocyanic acid **28**, or urea **26**. **28** reacts with ammonia in the presence of H_2O as a proton shuttling catalyst to produce **27**, which is further dehydrated in the presence of an NH_3 molecule acting as a catalyst, leading to the formation of **29** (shown in Figure 5 and the energy profile in Figure S12). There are several other possible pathways for the formation of **28**, which have been discussed in Figure S13, Supporting Information.

During the nanoreactor simulations, apart from simple and complex acyclic organic compounds, numerous cyclic com-

DOI: 10.1021/acscentsci.9b00520
ACS Cent. Sci. 2019, 5, 1532−1540

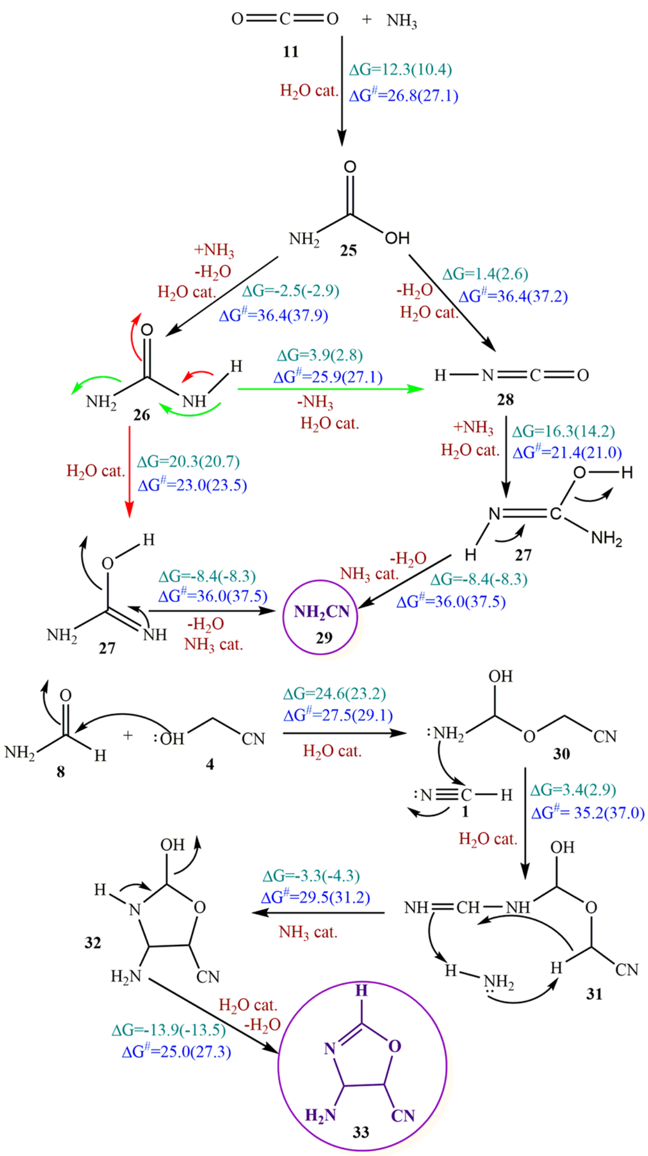


Figure 5. Formation of the target species: cyanamide and the oxazole derivative. The values are in kcal/mol.

pounds such as oxazole, imidazole, as well as isoxazole derivatives were also seen to have formed (see Figure S14, Supporting Information). Among these heterocyclic compounds, oxazole derivatives are among the more important, because 2-amino-oxazole is known to be an important precursor for RNA synthesis.[27,62,63] In our current work, we have shown a very feasible pathway for the formation of one of the oxazole derivatives **33** (encircled in violet in Figure 5) which is formed during the reaction between **8** with **4** (shown in Figure 5, with the energy profile in Figure S5, Supporting Information). For more information on the formation of cyclic products, please see Figures S7 and S15, Supporting Information.

Implications of the Current Work. HCN, as has been noted in the literature, may have occupied a "unique position in terrestrial pre-biological chemistry".[40] The current work shows that just the interaction between HCN and water as the starting reactants would have been sufficient to eventually lead to the precursors of RNA and proteins. This is significant because it shows that the reactions could have happened ubiquitously in the water bodies all over the Earth. What is also important is that all the conditions specified as (i−vi) earlier were adhered to during the simulations.

There are, however, important questions that need to be addressed. First, there is the issue of low concentrations of HCN in water in early Earth conditions, which would have reduced the concentration of the subsequently formed RNA and protein precursors. A resolution to this problem is suggested by a recent molecular dynamics report which indicates that in dilute systems the HCN concentration is an order of magnitude larger in the surface layer than in the bulk liquid phase.[71] Such HCN concentration effects at the surface of water bodies would have facilitated the chemistry described here. Furthermore, certain hydrothermal vents at the bottom of the ocean are in the vicinity of cold seawater, as well as ice. Recent reports suggest that HCN could be stabilized and concentrated at water−ice interfaces as well.[72,73] Water containing this more concentrated HCN could then have seeped into the hydrothermal vents in the vicinity of the cold water−ice and undergone hydrolysis at higher temperatures inside the vents.

The other issue is with regard to the eventual products of the HCN hydrolysis. Our results show that the formation of precursor molecules of RNA and protein would have been feasible, but the question then is would these precursors have been formed in sufficient concentrations to then react with each other, in order to lead to greater complexity? One solution that can be provided to this problem is to invoke the idea of a "warm little pond" that had been suggested by Charles Darwin in 1871,[74] i.e., to consider shallow ponds, lakes, estuaries, or tidal lagoons in prebiotic times that would have had temperatures of about 100.0 °C. The reactions discussed here could have happened in such water bodies, and then evaporation of the water would have led to increased concentration of the products formed. Such a scenario would have led to greater interactions between the molecules formed, and thus, to more complex molecules. It is also possible that such precursors could have seeped out with water from hydrothermal vents and been concentrated at water−ice interfaces in the vicinity of the vents, which would then have allowed subsequent, more complex molecules to have emerged.

Another clarification that should be made is with regard to the specified condition (iii) in the Introduction, regarding the need to have chemical reactions occurring without the need for metal catalysts. This condition increases the probability of the chemical reactions taking place all over the Earth's oceans, and not just in the few, select regions of the Earth where metal based catalysts were available. However, in the regions where metal based catalysts were available in early Earth, their presence would have been beneficial and accelerated the formation of the precursors for RNA and proteins.

The other salient points gleaned from the AINR studies are as follows:

(i) As the pathways found for sugar formation (see Figure 4B) indicate, lower barriers have been obtained for chemical conversions when an pathway alternative to reduction was found (see the respective steps in the reduction pathways, shown in Figure S9, Supporting Information). This corroborates experimental observations that indicate that reduction was generally avoided in prebiotic chemistry.[75]

DOI: 10.1021/acscentsci.9b00520
ACS Cent. Sci. 2019, 5, 1532−1540

(ii) The formation of low valent species such as carbenes (**17**, **12**, **14**, **15**, **16**, NH_2COH) is an important reason why most of the barriers for the mechanistic pathways discovered through the AINR approach were seen to be reasonable to low. This interesting fact echoes previous hypotheses that low valent main group compounds are important intermediates in mechanistic cycles.[70,76,77]

(iii) The AINR was seen to exploit H_2O or NH_3 molecules as proton shuttling catalysts in most of the elementary reaction steps. This, again, has relevance in the context of recent reports,[50,78,79] suggesting that a lot of biology occurs with the mediation of H_2O as a proton shuttling catalyst. Moreover, the role of NH_3 as a proton shuttling agent has also been explored in the literature.[80−83]

■ CONCLUSION

The current work shows that interaction between only two different molecules—HCN and H_2O—would have been sufficient to give rise to most of the important precursors to RNA and proteins in prebiotic times. Taking advantage of the recently developed AINR method,[50] which has allowed us to discover new reaction pathways, we have shown that cyanamide, glycoladehyde, oxazole derivative, and glycine all could have been formed from only a single carbon and nitrogen source molecule: HCN and a single oxygen source molecule: H_2O, at temperatures of about 80.0−100.0 °C. Pathways that were found to be feasible were seen to avoid the reduction step, corroborating previous experimental reports.[75] Most of the steps of the discovered mechanistic routes have barriers that are low to moderate, with only a few higher barriers of ∼40.0 kcal/mol, which suggests that the reactions could have occurred without the mediation of metal catalysts and through the aid of thermochemistry alone. This insight is valuable because it helps to explain how the reactions could have taken place in the absence of photochemical activity on the surface of Earth's oceans. Furthermore, the RNA and protein precursor molecules were obtained during the simulations in "one-pot", i.e., during a single simulation in the AINR.

These findings make it possible to imagine that the molecules necessary for building larger, more complex entities such as RNA and proteins could have existed and interacted together in at least some of the water bodies present in early Earth. The current work thus indicates that HCN and H_2O could have been the Adam and Eve of chemical evolution—the source of the precursor molecules that formed the basis of life on Earth.

■ COMPUTATIONAL METHODS

Ab Initio Molecular Dynamics (AIMD) Simulations. The nanoreactor AIMD simulations were performed with the TeraChem 1.9 software package[84−90] using the Hartree−Fock (HF)[91] electronic wave function and the 3-21g(d) Gaussian basis set,[92] to calculate the Born−Oppenheimer potential energy surface. This method has been implemented in TeraChem by Martinez and co-workers.[50] This approach was deemed acceptable because the HF method is well-known for predicting chemically reasonable structures.[93] Also, it should be noted HF was not employed to determine barrier heights and reaction rates: its only role was in the discovery process. This was also the approach employed by Martinez and co-workers in their original AINR paper (employing HF/3-21g), where they replicated the results obtained from the Urey−Miller experiment, as well as from the interaction of acetylene molecules.[50] We note here that we did also attempt discovery in the AINR simulations with DFT using the B3LYP density functional and the 3-21g(d) basis set and did find the preliminary intermediates (formamide, formic acid, formaldehyde and others) through this approach as with HF/3-21g(d). However, this was at greater computational expense and did not appear to give different results from the HF/3-21g(d) approach. Hence, we have limited the discovery process to HF/3-21g(d) in the AINR simulations.

The AINR simulation results that have been discussed here pertain to the case in which 16 H_2O and 15 HCN molecules were taken together in a spherical box of radii 10.0 and 3.5 Å (the system alternated between the two radii, in order for the collisions to take place—see original paper by Martinez and co-workers[50]). This system was allowed to evolve for 750 ps and generated the intermediates and reaction pathways that have been discussed. Additionally, we have also performed several AINR simulations where we varied different parameters, such as (i) the ratio of the reactant species, (ii) the total number of molecules taken in the simulation box, (iii) the spherical boundary conditions, (iv) the temperature, and (v) the total time of the AIMD simulations. The results obtained by changing the parameters (i−v) have been discussed in the Computational Details section in the Supporting Information. In general, they indicate that while most of the intermediates were discovered by varying (i) to (v), the most comprehensive results were obtained from the simulation case described in the manuscript: taking 16 H_2O and 15 HCN molecules in a spherical box of radii 10.0 and 3.5 Å. Moreover, for this case, multiple simulations were also performed from the same initial configuration and were seen to give rise to all the desired intermediates and products (following the same mechanistic pathways), although the time of formation of these species during the simulations was seen to change from simulation to simulation.

Newton's equations of motion were calculated using Langevin dynamics with an equilibrium temperature of 2000.0 K (also the starting temperature of the dynamics). We have used this high temperature in order to increase the average kinetic energy of the reactant molecules and for faster dynamics, allowing the overcoming of noncovalent interactions without the breaking of covalent bonds. This, too, follows the example of the work with the AINR done by Martinez and co-workers.[50] The nanoreactor simulations employ a piston to accelerate the reaction rate. We have employed the augmented direct inversion in the iterative subspace (ADIIS) algorithm[94] available in TeraChem as an alternative tool for self-consistent field calculations at each AIMD step in which the default DIIS algorithm[95] failed to converge. Spherical boundary conditions were applied to prevent the molecules from flying away, a phenomenon known as the "evaporation" event. For further information on the spherical boundary conditions, analysis of simulation trajectories by NetworkX, Numpy, and Graphviz Python libraries, please see the Supporting Information.

The mechanistic pathways obtained from the AINR simulations were then analyzed as follows: (i) all the reactant, intermediate, and transition state structures were optimized with high level density functional theory (DFT) calculations, using the Turbomole 7.0 software package at the B3LYP-D3/TZVP+COSMO(ε = 80.0) level of theory, (ii) single point calculations were then done at both the coupled cluster

DOI: 10.1021/acscentsci.9b00520
ACS Cent. Sci. 2019, 5, 1532−1540

(RICC2) as well as the Møller–Plesset second order perturbation (RIMP2) levels of theory in order to obtain the electronic energies. Hence, the calculations have been done at the B3LYP-D3/TZVP+COSMO(ε = 80.0)//RI-CC2/TZVP+COSMO(ε = 80.0) and the B3LYP-D3/TZVP+COSMO(ε = 80.0)//RI-MP2/TZVP+COSMO(ε = 80.0) (values shown in parentheses in the free energy profiles) levels of theory. Further, entropic and internal energy contributions were determined by frequency calculations at the B3LYP-D3/TZVP+COSMO(ε = 80.0) level of theory. Volume corrections were also included for the translational entropy term. Moreover, in addition to these calculations done with Turbomole 7.0, Gaussian 09 was also employed to obtain the free energy profiles, at the B3LYP-D2/6-311++g(d,p)+PCM(ε = 80.0) and the M06-2X/6-311++g(d,p)+PCM(ε = 80.0) levels of theory. The free energy profiles obtained from the Gaussian 09 calculations are provided in the Supporting Information and were seen to be similar to the profiles obtained by employing Turbomole 7.0. More details of the static DFT and QM calculations, along with all the relevant references, are provided in the Supporting Information.

Safety Statement. No unexpected or unusually high safety hazards were encountered.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acscentsci.9b00520.

Computational details, alternative mechanistic pathways, Cartesian coordinates of all the transition states (PDF)

Movie (MP4)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: k.vanka@ncl.res.in.

ORCID

Kumar Vanka: 0000-0001-7301-7573

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors acknowledge the Centre of Excellence in Scientific Computing (COESC), NCL, Pune, for providing computational facilities. K.V. is grateful to the Department of Science and Technology (DST) (EMR/2014/000013) for providing financial assistance. T.D. and S.G. thank Council of Scientific and Industrial Research (CSIR) for providing Research Fellowship.

■ REFERENCES

(1) Gánti, T. *The Principles of Life*; Oxford Univsersity Press: Oxford UK, 2003.
(2) Dyson, F. *Origins of Lfe*; Cambridge University Press: Cambridge UK, 1999.
(3) Sutherland, D. Studies on the origin of life — the end of the beginning. *Nat. Rev.* **2017**, *1*, 1–7.
(4) Schopf, J. W. *Life's Origin: The Beginnings of Biological Evolution*; University of California Press: Oakland, CA, 2002.
(5) Steel, M.; Penny, D. Origins of life: common ancestry put to the test. *Nature* **2010**, *465*, 168–169.
(6) Oparin, A. I. *The Origin of Life*; World Publishing: Cleveland, 2003.
(7) Courier, D.; Oro, J. *Historical Understanding of Life's Beginnings*; World Publishing: Cleveland, 2002.
(8) Bernal, J. D. *The Origin of Life*; World Publishing: Cleveland, 1967.
(9) Sutherland, J. D. The origin of life—out of the blue. *Angew. Chem., Int. Ed.* **2016**, *55*, 104–121.
(10) Bracher, P. J. Origin of life: primordial soup that cooks itself. *Nat. Chem.* **2015**, *7*, 273–274.
(11) Pascal, R.; Pross, A.; Sutherland, J. D. Towards an evolutionary theory of the origin of life based on kinetics and thermodynamics. *Open Biol.* **2013**, *3*, 130156–130164.
(12) Bada, J. L. How life began on earth: a status report. *Earth Planet. Sci. Lett.* **2004**, *226*, 1–15.
(13) Bada, J. L. New insights into prebiotic chemistry from Stanley Miller's spark discharge experiment. *Chem. Soc. Rev.* **2013**, *42*, 2186–2196.
(14) Woese, C. *The Genetic Code*; Harper & Row, 1967; pp 179–195.
(15) Crick, F. H. C. The origin of the genetic code. *J. Mol. Biol.* **1968**, *38*, 367–379.
(16) Orgel, L. E. Evolution of the genetic apparatus. *J. Mol. Biol.* **1968**, *38*, 381–393.
(17) Joyce, G. F. The antiquity of RNA-based evolution. *Nature* **2002**, *418*, 214–221.
(18) Joyce, G. F.; Orgel, L. E. In *The RNA World*; Gesteland, R. F.; Cech, T. R.; Atkins, J. F., Eds.; Cold Spring Harbor Laboratory Press: New York US, 2006; pp 23–56.
(19) Orgel, L. E. Prebiotic chemistry and the origin of the RNA world. *Crit. Rev. Biochem. Mol. Biol.* **2004**, *39*, 99–123.
(20) Schmidt, S.; Sunyaev, S.; Bork, P.; Dandekar, T. Metabolites: a helping hand for pathway evolution. *Trends Biochem. Sci.* **2003**, *28*, 336–341.
(21) Caetano-Anolles, G.; Kim, H. S.; Mittenthal, J. A. The origin of modern metabolic networks inferred from phylogenomic analysis of protein architecture. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 9358–9363.
(22) Copley, R. R.; Bork, P. Homology among $(\beta\alpha)_8$ barrels: implications for the evolution of metabolic pathways. *J. Mol. Biol.* **2000**, *303*, 627–640.
(23) Danchin, A. From chemical metabolism to life: the origin of the genetic coding process. *Beilstein J. Org. Chem.* **2017**, *13*, 1119–1135.
(24) Ferris, J. P.; Hill, A. R.; Liu, R., Jr; Orgel, L. E. Synthesis of long prebiotic oligomers on mineral surfaces. *Nature* **1996**, *381*, 59–61.
(25) Verlander, M. S.; Lohrmann, R.; Orgel, L. E. Catalysts for the self-polymerization of adenosine cyclic 2′,3′-phosphate. *J. Mol. Evol.* **1973**, *2*, 303–316.
(26) Szostak, J. W. System chemistry on early earth. *Nature* **2009**, *459*, 171–172.
(27) Powner, M. W.; Gerland, B.; Sutherland, J. D. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* **2009**, *459*, 239–242.
(28) Ritson, D. J.; Sutherland, J. D. Synthesis of aldehydic ribonucleotide and amino acid precursors by photoredox chemistry. *Angew. Chem., Int. Ed.* **2013**, *52*, 5845–5847.
(29) Bowler, F. R.; Chan, C. K.; Duffy, C. D.; Gerland, B.; Islam, S.; Powner, M. W.; Sutherland, J. D.; Xu, J. Prebiotically plausible oligoribonucleotide ligation facilitated by chemoselective acetylation. *Nat. Chem.* **2013**, *5*, 383–389.
(30) Patel, B. H.; Percivalle, C.; Ritson, D. J.; Duffy, C. D.; Sutherland, J. D. Common origins of RNA, protein and lipid precursors in a cyanosulfidic protometabolism. *Nat. Chem.* **2015**, *7*, 301–307.
(31) Ritson, D. J.; Battilocchio, C.; Ley, S. V.; Sutherland, J. D. Mimicking the surface and prebiotic chemistry of early Earth using flow chemistry. *Nat. Commun.* **2018**, *9*, 1821–1830.
(32) Anastasi, C.; Buchet, F. F.; Crowe, M. A.; Helliwell, M.; Raftery, J.; Sutherland, J. D. The search for a potentially prebiotic synthesis of

nucleotides via arabinose-3-phosphate and its cyanamide derivative. *Chem. - Eur. J.* **2008**, *14*, 2375−2388.
(33) Sanchez, R. A.; Ferris, J. P.; Orgel, L. E. Cyanoacetylene in prebiotic synthesis. *Science* **1966**, *154*, 784−785.
(34) Ferris, J. P.; Hagan, W. J., Jr. HCN and chemical evolution: the possible role of cyano compounds in prebiotic synthesis. *Tetrahedron* **1984**, *40*, 1093−1120.
(35) Schlesinger, G.; Miller, S. L. Prebiotic synthesis in atmospheres containing CH_4, CO and CO2. II Hydrogen cyanide, formaldehyde and ammonia. *J. Mol. Evol.* **1983**, *19*, 383−390.
(36) Miller, S. L. A production of amino acids under possible primitive earth conditions. *Science* **1953**, *117*, 528−529.
(37) Bada, J. L.; Cleaves, H. J. *Ab initio* simulations and the Miller prebiotic synthesis experiment. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, E342.
(38) Arney, N. G; Domagal-Goldman, D. S; Meadows, S. V. Organic haze as a biosignature in anoxic earth-like atmospheres. *Earth and Planetary Astrophysics* **2017**, arXiv:1711.01675.
(39) Kasting, J. F.; Ackerman, T. P. Climatic consequences of very high carbon dioxide levels in the earth's early atmosphere. *Science* **1986**, *234*, 1383−1385.
(40) Sanchez, R. A.; Ferris, J. P.; Orgel, L. E. Synthesis of purine precursors and amino acids from aqueous hydrogen cyanide. *J. Mol. Biol.* **1967**, *30*, 223−253.
(41) Aubrey, A. D.; Cleaves, H. J.; Bada, J. L. The role of submarine hydrothermal systems in the synthesis of amino acids. *Origins Life Evol. Biospheres* **2009**, *39*, 91−108.
(42) White, R. H. Hydrolytic stability of biomolecules at high temperatures and its implication for life at 250°C. *Nature* **1984**, *310*, 430−432.
(43) Miller, S. L.; Bada, J. L. Submarine hot springs and the origin of life. *Nature* **1988**, *334*, 609−611.
(44) Szori, M.; Jojart, B.; Izsak, R.; Szori, K.; Csizmadia, I. G.; Viskolcz, B. Chemical evolution of biochemical building blocks. Can thermodynamics explain the accumulation of glycine in the prebiotic ocean? *Phys. Chem. Chem. Phys.* **2011**, *13*, 7449−7458.
(45) Florian, J.; Warshel, A. Phosphate ester hydrolysis associative in aqueous solution: Associative versus dissociative mechanisms. *J. Phys. Chem. B* **1998**, *102*, 719−734.
(46) Song, J. L.; Wang, T.; Zhang, X.; Chung, W. L.; Wu, D. Y. A Combined DFT/IM-MS Study on the reaction mechanism of cationic Ru(II)-catalyzed hydroboration of alkynes. *ACS Catal.* **2017**, *7*, 1361−1368.
(47) Lee, B. T.; Mckee, L. M. Mechanistic study of $LiNH_2BH_3$ formation from $(LiH)_4$ + $NH_3BH_3$and subsequent dehydrogenation. *Inorg. Chem.* **2009**, *48*, 7564−7575.
(48) Nguyen, T. M.; Nguyen, S. V.; Matus, H. M.; Gopakumar, G.; Dixon, A. D. Molecular mechanism for H_2 release from BH_3NH_3, including the catalytic role of the Lewis acid BH_3. *J. Phys. Chem. A* **2007**, *111*, 679−690.
(49) Yamakawa, M.; Ito, H.; Noyori, R. The metal−ligand bifunctional catalysis: A theoretical study on the ruthenium(II)-catalyzed hydrogen transfer between alcohols and carbonyl compounds. *J. Am. Chem. Soc.* **2000**, *122*, 1466−1478.
(50) Wang, L. P.; Titov, A.; McGibbon, R.; Liu, F.; Pande, V. S.; Martinez, T. J. Discovering chemistry with an ab initio nanoreactor. *Nat. Chem.* **2014**, *6*, 1044−1048.
(51) Zimmerman, P. M. Automated discovery of chemically reasonable elementary reaction steps. *J. Comput. Chem.* **2013**, *34*, 1385−1392.
(52) Rappoport, D.; Galvin, C. J.; Zubarev, D. Y.; Aspuru-Guzik, A. Complex chemical reaction networks from heuristics-aided quantum chemistry. *J. Chem. Theory Comput.* **2014**, *10*, 897−907.
(53) Miller, S. L.; Urey, H. C. Organic compound synthesis on the primitive earth. *Science* **1959**, *130*, 245−251.
(54) Turner, B. E.; Liszt, H. S.; Kaifu, N.; Kisliakov, A. G. Microwave detection of interstellar cyanamide. *Astrophys. J.* **1975**, *201*, L149−L152.
(55) Lohrmann, R. Formation of urea and guanidine by irradiation of ammonium cyanide. *J. Mol. Evol.* **1972**, *1*, 263−269.
(56) Decker, P.; Schweer, H.; Pohlamnn, R. Bioids: X. Identification of formose sugars, presumable prebiotic metabolites, using capillary gas chromatography/gas chromatography−mass spectrometry of n-butoxime trifluoroacetates on OV-225. *J. Chromatogr. A* **1982**, *244*, 281−291.
(57) Ricardo, A.; Carrigan, M. A.; Olcott, A. N.; Benner, S. A. Borate minerals stabilize ribose. *Science* **2004**, *303*, 196.
(58) Serianni, A. S.; Clark, E. L.; Barker, R. Carbon-13-enriched carbohydrates. Preparation of erythrose, threose, glyceraldehyde, and glycolaldehyde with ^{13}C-enrichment in various carbon atoms. *Carbohydr. Res.* **1979**, *72*, 79−91.
(59) Fischer, E. Reduction von säuren der Zuckergruppe. *Ber. Dtsch. Chem. Ges.* **1889**, *22*, 2204−2205.
(60) Ritson, D.; Sutherland, J. D. Prebiotic synthesis of simple sugars by photoredox systems chemistry. *Nat. Chem.* **2012**, *4*, 895−899.
(61) Benner, S. A.; Kim, H. J.; Carrigan, M. A. Asphalt, water, and the prebiotic synthesis of ribose, ribonucleosides, and RNA. *Acc. Chem. Res.* **2012**, *45*, 2025−2034.
(62) Cockerill, A. F.; Deacon, A.; Harrison, R. G.; Osborne, D. J.; Prime, D. M.; Ross, W. J.; Todd, A.; Verge, J. P. An improved synthesis of 2-amino-1,3-oxazoles under basic catalysis. *Synthesis* **1976**, *1976*, 591−593.
(63) Eschenmoser, A.; Loewenthal, E. Chemistry of potentially prebiological natural products. *Chem. Soc. Rev.* **1992**, *21*, 1−16.
(64) Strecker, A. Ueberein enneuenaus aldehyd-ammoniak und blausäureents tehenden Körpe. *Liebigs Ann. Chem.* **1854**, *91*, 349−351.
(65) Bar-Nun, A.; Bar-Nun, N.; Bauer, S. H.; Sagan, C. Shock synthesis of amino acids in simulated primitive environments. *Science* **1970**, *168*, 470−473.
(66) Matthews, C. N.; Moser, R. E. Peptide synthesis from hydrogen cyanide and water. *Nature* **1967**, *215*, 1230−1234.
(67) Goldman, N.; Reed, E. J.; Fried, L. E.; William Kuo, I.-F.; Maiti, A. Synthesis of glycine containing complexes in impact of comets on early earth. *Nat. Chem.* **2010**, *2* (11), 949−954.
(68) Saitta, A. M.; Saija, F. Miller experiments in atomistic computer simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 13768−13773.
(69) Thaddeus, P. The prebiotic molecules observed in the interstellar gas. *Philos. Trans. R. Soc., B* **2006**, *361*, 1681−1687.
(70) Eckhardt, A. K.; Linden, M. M.; Wende, R. C.; Bernhardt, B.; Schreiner, P. R. Gas-phase sugar formation using hydroxymethylene as the reactive formaldehyde isomer. *Nat. Chem.* **2018**, *10*, 1141−1147.
(71) Fabian, B.; Szori, M.; Jedlovszky, P. Floating patches of HCN at the surface of their aqueous solutions − can they make "HCN World" plausible? *J. Phys. Chem. C* **2014**, *118*, 21469−21482.
(72) Szori, M.; Jedlovszky, P. Adsorption of HCN at the Surface of Ice: A grand canonical Monte Carlo simulation study. *J. Phys. Chem. C* **2014**, *118*, 3599−3609.
(73) Menor-Salván, C.; Marín-Yaseli, R. M. Prebiotic chemistry in eutectic solutions at the water−ice matrix. *Chem. Soc. Rev.* **2012**, *41*, 5404−5415.
(74) Ball, P. *Elegant Solutions: Ten Beautiful Experiments in Chemistry*; The Royal Society of Chemistry: Cambridge UK, 2005; pp , 1−205.
(75) Danger, G.; Plasson, R.; Pascal, R. Pathways for the formation and evolution of peptides in prebiotic environments. *Chem. Soc. Rev.* **2012**, *41*, 5416−5429.
(76) Mandal, S. K.; Roesky, H. W. Group 14 hydrides with low valent elements for activation of small molecules. *Acc. Chem. Res.* **2012**, *45*, 298−307.
(77) Nesterov, V.; Reiter, D.; Bag, P.; Frisch, P.; Holzner, R.; Porzelt, A.; Inoue, S. NHCs in main group chemistry. *Chem. Rev.* **2018**, *118*, 9678−9842.
(78) Mikulski, R.; West, D.; Sippel, K. H.; Avvaru, B. S.; Aggarwal, M.; Tu, C.; McKenna, R.; Silverman, D. N. Water network in fast

DOI: 10.1021/acscentsci.9b00520
ACS Cent. Sci. 2019, 5, 1532−1540

proton transfer during catalysis by human carbonic anhydrase II. *Biochemistry* **2013**, *52*, 125−131.
(79) De Vivo, M.; Ensing, B.; Klein, M. L. Computational study of phosphatase activity in soluble epoxide hydrolase: High efficiency through a water bridge mediated proton shuttle. *J. Am. Chem. Soc.* **2005**, *127*, 11226−11227.
(80) Pal, A.; Vanka, K. Small molecule activation by constrained phosphorous compounds: Insight from theory. *Inorg. Chem.* **2016**, *55*, 558−565.
(81) Cord-Ruwisch, R.; Law, Y.; Cheng, K. Y. Ammonium as a sustainable proton shuttle in bio electrochemical systems. *Bioresour. Technol.* **2011**, *102*, 9691−9696.
(82) Meuwly, M.; Karplus, M. Simulation of proton transfer along ammonia wires: An "*ab initio*" and semiempirical density functional comparison of potentials and classical molecular dynamics. *J. Chem. Phys.* **2002**, *116*, 2572−2585.
(83) Jaroszewski, L.; Lesyng, B.; Tanner, J. J.; McCammon, J.A. *Ab initio* study of proton transfer in $[H_3N\text{-}H\text{-}NH_3]^+$ and $[H_3N\text{-}H\text{-}H_2O]^+$. *Chem. Phys. Lett.* **1990**, *175*, 282−288.
(84) Ufimtsev, I. S.; Martinez, T. J. Quantum chemistry on graphical processing units. 3. Analytical energy gradients, geometry optimization, and first principles molecular dynamics. *J. Chem. Theory Comput.* **2009**, *5*, 2619−2628.
(85) Ufimtsev, I. S.; Luehr, N.; Martinez, T. J. Charge transfer and polarization in solvated proteins from *ab initio* molecular dynamics. *J. Phys. Chem. Lett.* **2011**, *2*, 1789−1793.
(86) Isborn, C. M.; Luehr, N.; Ufimtsev, I. S.; Martinez, T. J. Excited-state electronic structure with configuration interaction singles and Tamm−Dancoff time dependent density functional theory on graphical processing units. *J. Chem. Theory Comput.* **2011**, *7*, 1814−1823.
(87) Titov, A. V.; Ufimtsev, I. S.; Luehr, N.; Martinez, T. J. Generating efficient quantum chemistry codes for novel architectures. *J. Chem. Theory Comput.* **2013**, *9*, 213−221.
(88) Ufimtsev, I. S.; Martinez, T. J. Graphical processing units for quantum chemistry. *Comput. Sci. Eng.* **2008**, *10*, 26−34.
(89) Ufimtsev, I. S.; Martinez, T. J. Quantum chemistry on graphical processing units. 1. Strategies for two-electron integral evaluation. *J. Chem. Theory Comput.* **2008**, *4*, 222−231.
(90) Ufimtsev, I. S.; Martinez, T. J. Quantum chemistry on graphical processing units. 2. Direct self-consistent-field implementation. *J. Chem. Theory Comput.* **2009**, *5*, 1004−1015.
(91) Froese, F. C. General Hartree-Fock program. *Comput. Phys. Commun.* **1987**, *43*, 355−365.
(92) Binkley, J. S.; Pople, J. A.; Hehre, W. J. Self-Consistent molecular orbital methods. 21. Small split-valence basis sets for first-row elements. *J. Am. Chem. Soc.* **1980**, *102*, 939−947.
(93) Feller, D.; Peterson, K. A. An examination of intrinsic errors in electronic structure methods using the Environmental Molecular Sciences Laboratory computational results database and the Gaussian-2 set. *J. Chem. Phys.* **1998**, *108*, 154−176.
(94) Hu, X.; Yang, W. Accelerating self-consistent field convergence with the augmented Roothaan−Hall energy function. *J. Chem. Phys.* **2010**, *132*, 054109.
(95) Pulay, P. Convergence acceleration of iterative sequences—the case of SCF iteration. *Chem. Phys. Lett.* **1980**, *73*, 393−398.

DOI: 10.1021/acscentsci.9b00520
ACS Cent. Sci. 2019, 5, 1532−1540